DOCUMENT RESUME

ED 178 614 TM 009 975

AUTHOR Echternacht, Gary; Swinton, Spencer
TITLE Getting Straight: Everything You Always Wanted to Know about the Title I Regression Model and Curvilinearity.
PUB DATE Apr 79
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, CA, April 8-12, 1979) ; Figures marginally legible

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Gains; *Analysis of Covariance; Educational Assessment; Elementary Secondary Education; Evaluation Methods; Goodness of Fit; Graphs; Hypothesis Testing; Mathematical Models; *Program Evaluation; *Research Design; Research Problems; Scores
IDENTIFIERS Elementary Secondary Education Act Title I; Regression Effects; *RMC Models

ABSTRACT

Title I evaluations using the RMC Model C design depend for their interpretation on the assumption that the regression of posttest on pretest is linear across the cut score level when there is no treatment; but there are many instances where nonlinearities may occur. If one applies the analysis of covariance, or model C analysis, large errors may result. Various methods are suggested for model C users to deal with non-linearities. One method is Mosteller and Tukey's re-expression of posttest scores. Another method depends upon differential weighting of scores in different parts of the pretest distribution. A third method uses quadratic regression lines and extrapolation to obtain a "no-treatment expectation." Still another method is to assume parallel regression lines for the treatment and control groups. A number of different logistic curves may be tried if both floor and ceiling effects are present. Graphing the scatterplot of pretest versus posttest scores is also recommended. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178614

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

GETTING STRAIGHT:

Everything you always wanted to know about the Title I Regression Model and curvilinearity.

Presented at the Annual AERA Meeting 1979

San Francisco, CA

Gary Echternacht

and

Spencer Swinton

Educational Testing Service

Princeton, NJ

TM009 975

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gary Echternacht

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

INTRODUCTION

School districts receiving ESEA Title I funds are required to evaluate their projects using one of three evaluation models. One of these models, "model C," is a variation of the analysis of covariance and of Campbell and Stanley's regression and discontinuity design. Although relatively new to educational evaluation, the model was used as early as 1956 by Belsen to evaluate the effects of a television broadcast. Later the model was discussed in relation to the analysis of covariance by Cochran.

In the analysis of covariance, the treatment effect is indicated by $\bar{y}_T - \bar{y}_c - \beta\ (\bar{x}_T - \bar{x}_c)$. In the traditional analysis of covariance β is estimated from both the treatment and control group data. In Belsen's procedure, and in the Title I model, β is estimated from only the control group data.

In the Title I evaluation model, students are assigned to Title I based on the results of a pretest. Those scoring above a predetermined cut-score serve as a control group for those receiving the treatment. A posttest-on-pretest regression function is determined for the control group students. This function is also called the "no treatment expectation" function and is used to express the model C impact, which takes the form $\bar{y} - \bar{y}_c - \hat{\beta}_c\ (\bar{x}_T - \bar{x}_c)$. This is equivalent to extrapolating the regression function that is obtained for the control students to the area of the Title I students.

Implicit in the model C assumptions is that the regression function is linear when there is no treatment to begin with. Nonetheless, there

are many instances when nonlinearities may occur. For example, a ceiling effect on the posttest will result in a quadratic regression with a negative coefficient for the X^2 variable. Floor effects, differential skewing on pretest and posttests, and differential kurtosis on pretest and posttest may also lead to nonlinear posttest-on-pretest regressions. Indeed, even when pretest and posttest score distributions are identical, there is no guarantee that linear regressions will be obtained.

If one applies the analysis of covariance or model C analysis blindly to data that are distinctly non-linear, large errors may result. Large impacts can be made to look as though they were negative. Slight impacts can be made to look astoundingly large. In this display, we show various ways that have been suggested for dealing with non-linearities and give our suggestions for model C users. We attempt to show pitfalls in the analyses when nonlinearities occur. We welcome discussion of this preliminary work.

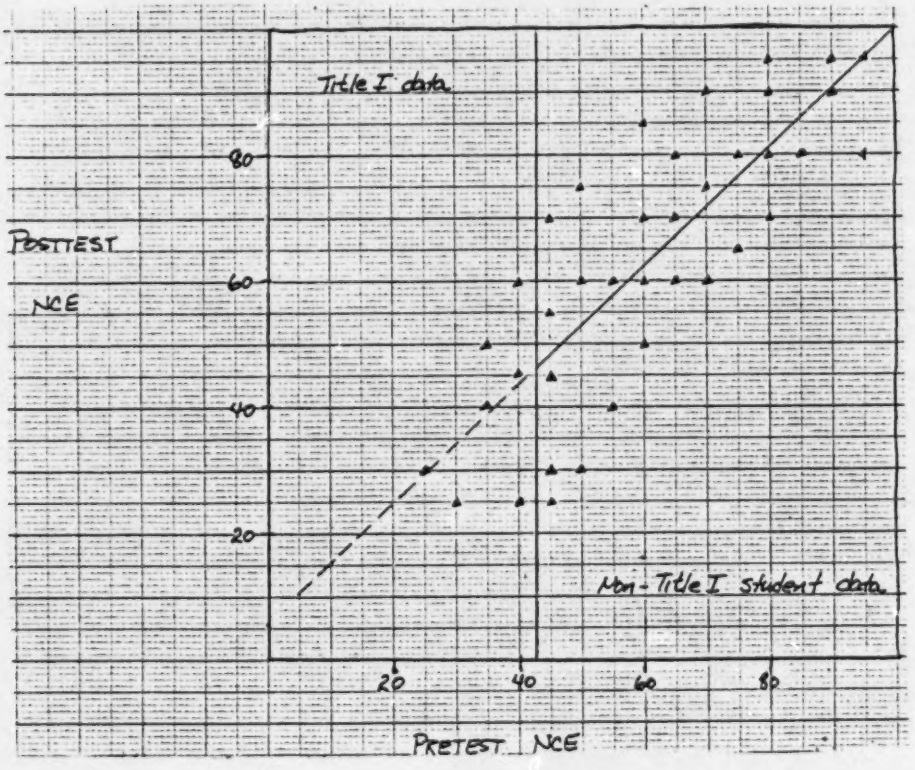
Re-expression

Although the computer is a basic tool of the research scientist, only rarely do school district evaluators have access to tools more powerful than a hand calculator. Nonetheless, technology has provided us impressive hand tools. For less than $50 one can buy a hand calculator that will produce regression analyses or correlations. The availability of such computing power should not blind us to the importance of graphing data and looking at the results, however.

In Mosteller and Tukey's book Data Analysis and Regression, a methodology for straightening curves and plots is presented. An example follows:

The Figure 1 below shows pretest and posttest data plotted for a hypothetical school. Students are selected for Title I based only on the pretest score. A straight line has been fit to the non-Title I student data and extrapolated to the Title I portion of the data to indicate the "no treatment expectation" for the model. The NCE gain is found to be -.2 NCE* score points.

FIGURE 1



*NCE's, or Normal Curve Equivalents, are a standardized score due to F. B. Davis. There are 99 equal NCE's from the first to the 99th percentile of a normal distribution. Thus an NCE is one-eleventh of the more familiar "stanine".

In observing the above plot, a slight bend in the data can be observed for the non-Title I students. We desire to straighten that plot.

Mosteller and Tukey's methodology consists of identifying three points and computing the slopes of the two lines connecting adjacent points. Ideally the slopes are equal for the two intervals. If they are not, the posttest scores can be transformed until the two intervals are equal.

In our example we chose five data points near the cut-score, three data points in the middle of the pretest score distribution, and three points at the high end of the distribution. We could have selected points in other ways. We chose the following points and averaged:

Low: ave $\{(45,25), (45,30), (45,45), (45,55), (45,70)\} = (45,45)$

Middle: ave $\{(70,60), (70,75), (70,90)\} = (70,75)$

High: ave $\{(90,95), (95,80), (95,95)\} = (93.3,90)$

The above three points, obtained by averaging, become the basis for re-expressing the posttest scores.

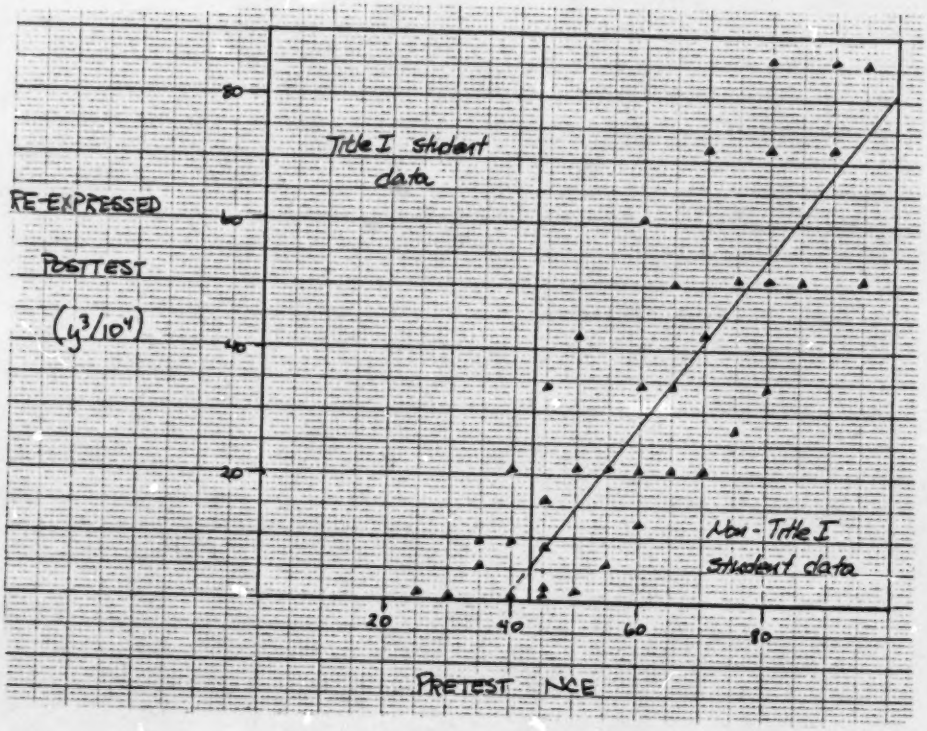
In straightening curves, we must go through the "ladder of re-expressions" (Mosteller & Tukey, p. 79) until we find a way to re-express (or transform) posttest scores so that the re-expressed posttest scores will bear a linear relationship to pretest scores from the non-Title I students.

Working through the ladder we obtain:

Expression	left interval slope	right interval slope	ratio
Y	1.20	.64	.54
$Y^2/10^2$	1.44	1.06	.74
$Y^3/10^4$	1.32	1.32	1.00

Since $Y^3/10^4$ results in equal right and left intervals, we use this to re-express the posttest scores in the form $Y^3/10^4$. Plots of the re-expressed scores appear below in Figure 2.

FIGURE 2



A straight line has been fit to the non-Title I students' data and extrapolated to the Title I student data. If NCE gains were calculated, the gain estimate would be 43.2 NCE points.

Certain conclusions and warnings can be seen in examining this example. They are:

1. The method is rough and as much an art as a science.

2. If there are few data points, selection of data points for averaging can be critical in the sense that different choices of re-expression will result. For example, had we chosen (90,90) and (90,95) as the two high points for averaging, we would have been led to re-express Y in terms of $Y^2/10^2$, yielding an estimated gain of 20.2 NCEs.

3. Sometimes when we re-express (as above) extrapolation goes below zero on the Y axis. This means that the expected NCE score is less than 1 and may require negative values on our scale.

4. In extrapolating the no treatment expectation, sometimes we come up with unreasonable expectations, as in the above example.

WEIGHTING SCORES

If a computer is available to the evaluator, more precise methods can be used to deal with non-linear regressions. Weighting the data would seem to be a logical procedure. Those who score high on pretest and posttest are less like those who score near the cut score, so we might want to give students who score near the cut-score more weight than those who pretest and posttest scores are high. Such a procedure would reduce the effect of any ceiling effects that we might find for the control group students.

In weighting, of course, the major problem is in selecting the weighting scheme to be used. But before we began to tackle this problem we investigated the likely efforts of any weighting scheme by choosing two weighting schemes and applying them to data that exhibited a marked ceiling effect, as in Figure 3.

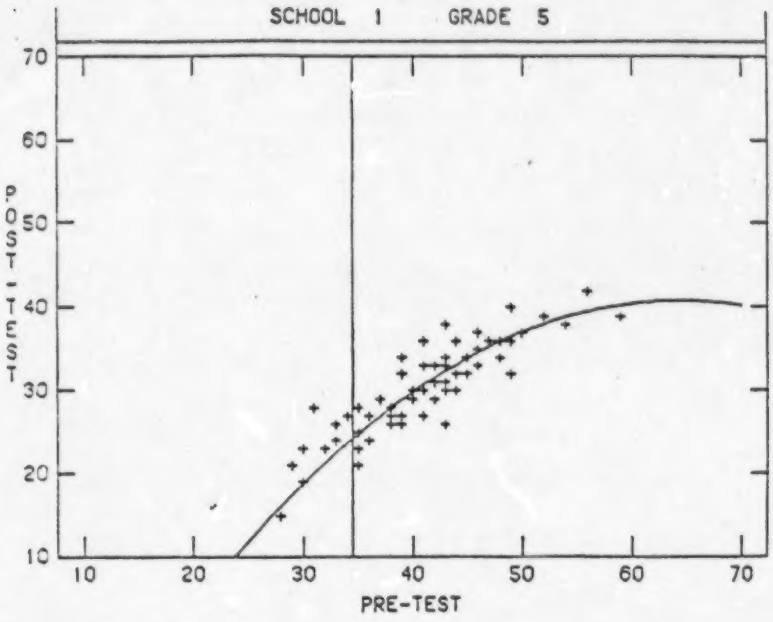
The two schemes we chose were:

$$W_{1i} = \frac{100}{100 + \tan^{-1}\left(\frac{(X - X_o)}{X_o - X_o}\right)}$$

and

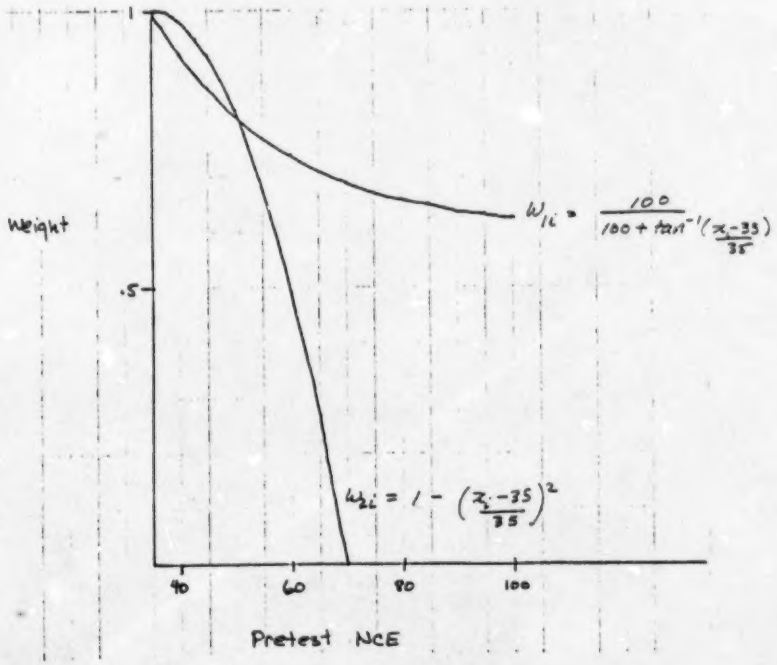
$$W_{2i} = 1 - (X - X_o)^2 / X_o^2$$

Figure 3



Where X_o represents the pretest cutscore. Graphs of the two functions appear below in Figure 4.

Figure 4



An example of one application of the two weighting schemes is presented below:

Independent variable	Regression weights	T-value (45 d.f.)	Multiple R
Pre	2.4179	2.9	.8479
Post	-0.0188	2.0	
Pre W_1	2.4182	2.9	.8478
Pre ^{2} W_1	-0.0188	2.0	
Pre W_2	2.4025	2.6	.8359
Pre ^{2} W_2	-0.0187	1.8	

From the above, one concludes that the unweighted and weighted (W_1) regressions are very similar. The only difference occurs in the coefficient of the linear term in the fourth decimal. The unweighted and W_2 weighting regressions are also very similar. So much so that the regression line indicated below for this data could be either the weighted or unweighted case. In any event, the moral is clear: with these typical Title I data, weighting produces essentially the same regression function as one would obtain with an unweighted fit.

GENERAL UNWEIGHTED MODELS

Again, assuming the evaluator has access to a computer, the following procedures might be considered. We illustrate each with three examples which use the same three data sets across methods.

QUADRATIC FITS

Although it generally does not deal with floor effects for the Title I students, one possible means of dealing with nonlinearities is to fit a quadratic regression to the control group data and extrapolate that regression to obtain a "no-treatment expectation" for the Title I group. We show three examples below of what might result. In two instances, figures 5 and 6, the data appear nonlinear; in one instance, figure 7, we have an essentially linear relationship. Although the "no-treatment expectations" in the cases of Figures 6 and 7 are reasonable, that pictured in Figure 5 is clearly absurd.

Figure 5

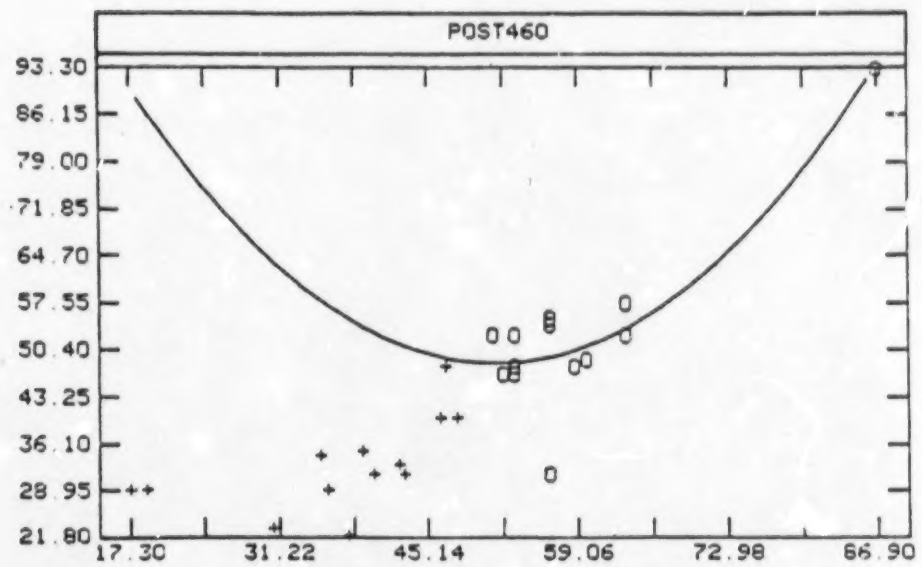


Figure 6

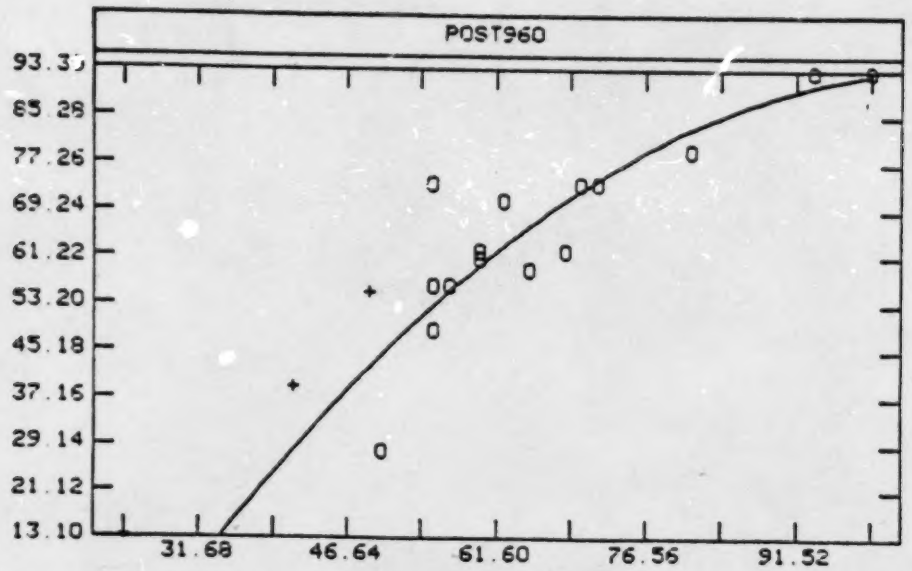
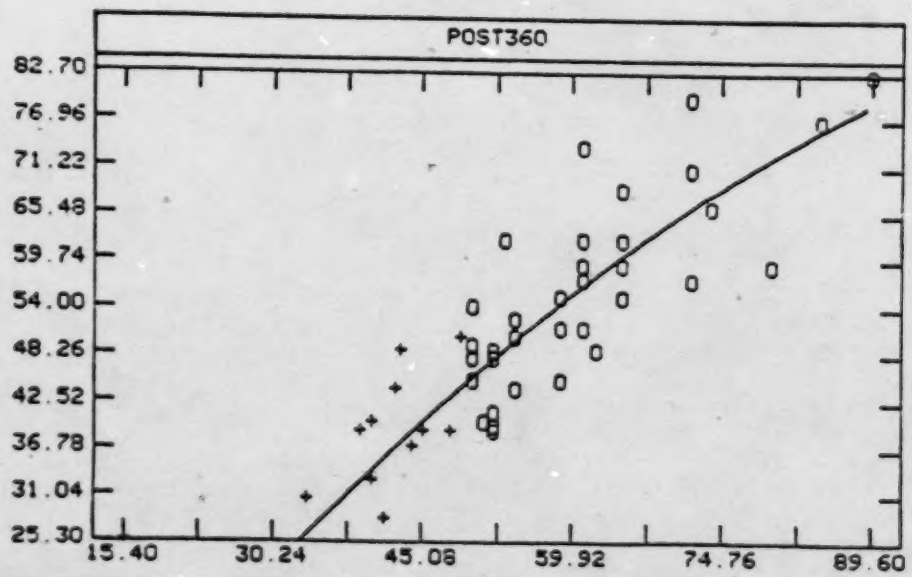


Figure 7



Conclusions about fitting quadratics:

1. When the data are nearly linear, one obtains a fit that looks linear.

2. When there are few data points, fitting quadratics can be dangerous.

3. Extrapolating quadratics can sometimes lead to ridiculous results.

4. Sometimes quadratic fits work well.

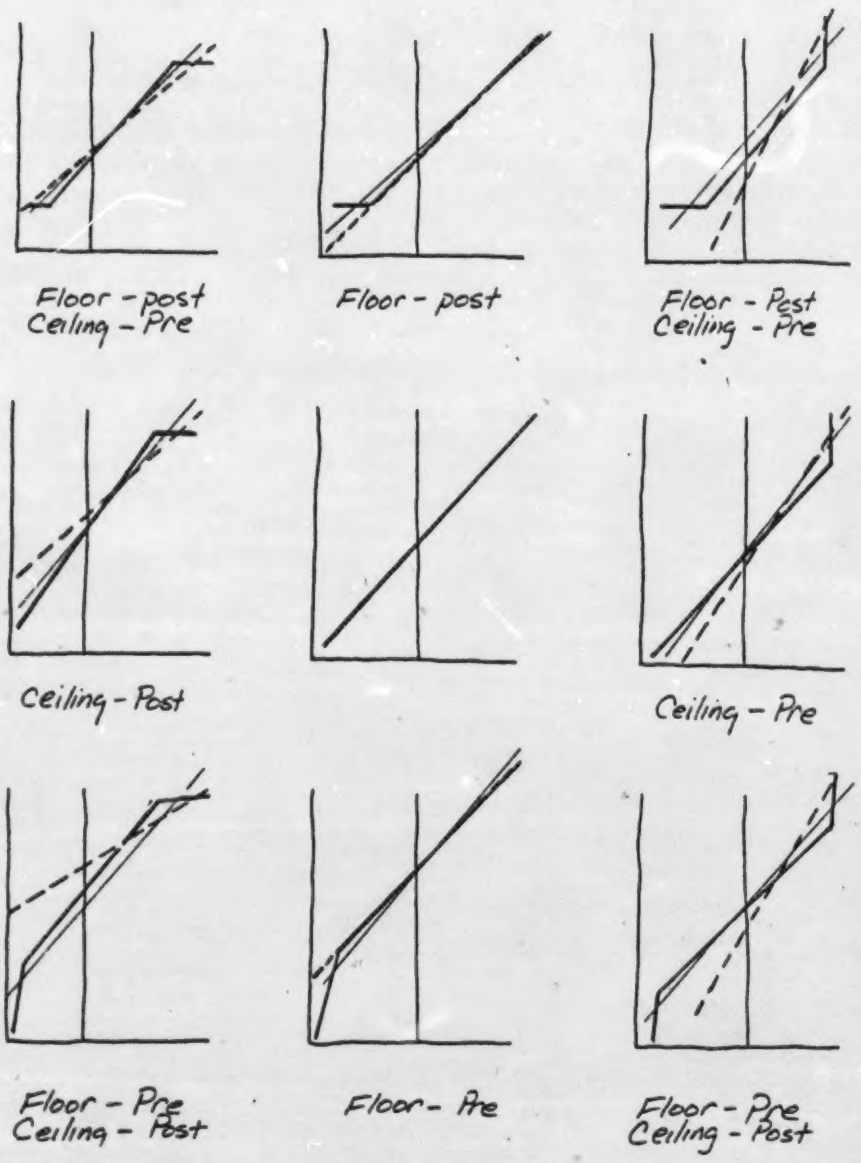
5. Quadratic fits do not take into account floor effects unless the effect reaches the control group.

PARALLEL LINES FIT

If one examines enough quadratic fits, one soon comes to the conclusion that "wild" estimates will sometimes occur with blind fitting and extrapolation of quadratics. Another possibility is to assume that the slope of the linear regression function will be the same for both treatment and control students. In this instance a treatment effect is defined as the distance between the two regression lines. This is an application of traditional analysis of covariance.

In examining how fitting parallel regression lines differs from fitting a linear regression to the control data and extrapolating, the following figure (8) is helpful.

FIGURE 8


Floor - post
Ceiling - Pre
Floor - post
Floor - Post
Ceiling - Pre
Ceiling - Post
Ceiling - Pre
Floor - Pre
Ceiling - Post
Floor - Pre
Floor - Pre
Ceiling - Post
LEGEND
Red
Blue
Black

The actual regression function is shown in black for various combinations of pretest and posttest floor and ceiling effects. The red line shows the regression that is fit for the control students and extrapolated for the treatment students. The blue line shows the parallel slope fit. In no cases do we show a treatment effect.

The figures above suggest that whenever ceiling effects are present, the parallel slopes fit is better than the model C procedure. Fitting parallel slopes may not be exact, but we are less likely to encounter "wild" results than when extrapolating.

Figures 9, 10, and 11 show parallel lines fit to the same data as was displayed in Figures 5,6, and 7. The slope may also be pooled across all schools, with a "dummy variable" constant introduced to allow for main effects affecting both treatment and control groups equally within a given school.

Figure 9

Parallel Line Fit

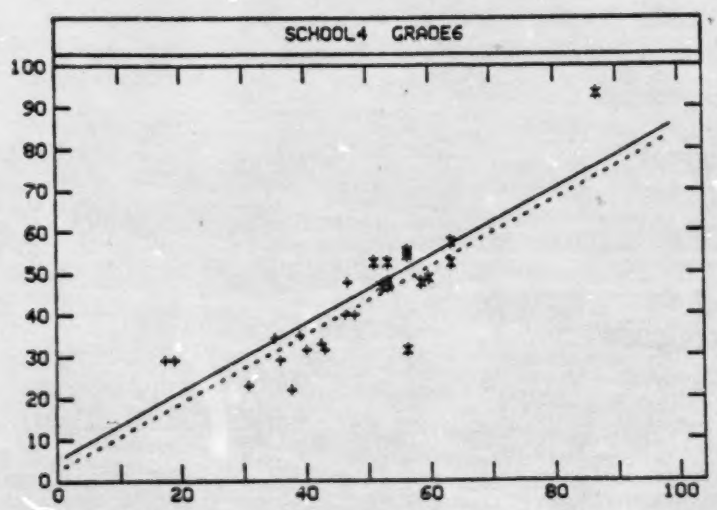


Figure 10

Parallel Line Fit

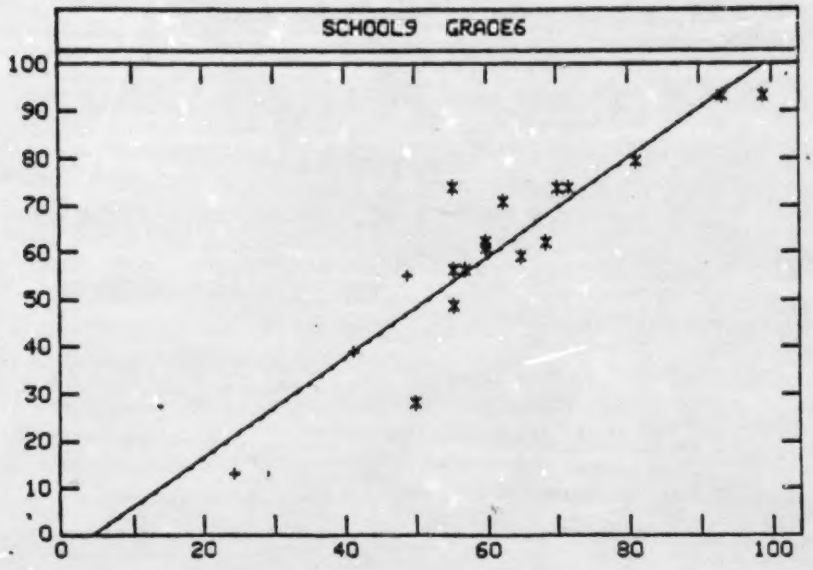
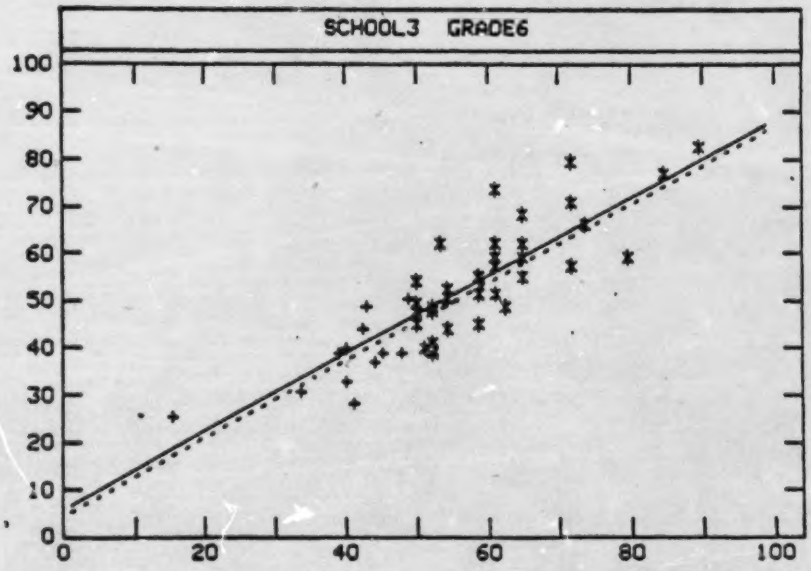


Figure 11

Parallel Line Fit



Conclusions about fitting parallel slopes:

1. One is less likely to obtain "wild" results from fitting parallel slopes than when extrapolating quadratics.
2. When floor effects only are encountered, extrapolation is slightly better than parallel lines.
3. If there is a treatment x pretest interaction, it will be confounded with the impact estimate resulting from fitting parallel lines.
4. If there is no treatment x pretest interaction, Cochran has shown that analysis of covariance or parallel lines fitting is better unless the number of control students is <u>much</u> larger than the number of treatment students.

LOGISTIC FITS

If both floor and ceiling effects are present, a logistic function would perhaps best describe Pretest-Posttest behavior in the absence of treatment. We would need a derivative-free regression routine since we want to estimate a treatment effect, which is indicated by a discontinuity at the cut score. The function should be approximately linear in cases when the data warrant such a fit. In fitting a logistic function, the treatment point-effect is defined as the distance between the curves at the cut-score, or alternatively (<u>not</u> the same thing) as in model C, the average distance between each treatment child's observed posttest score and an extrapolated logistic with P2 (the treatment effect) set to zero. In the function employed, P1 is a vertical translation parameter, δ a coded dummy variable indicating treatment/non-treatment, P3 a scaling

parameter which, if large, makes the function linear in the range of interest, P4 a slope or "discrimination" parameter, and P5 a horizontal translation parameter, analogous to difficulty in the more familiar test theoretical applications of this function. Figure 12 shows a dozen forms which the data may lead this flexible function to assume. Figures 13-15 show the forms it in fact did assume with our now-familiar three schools.

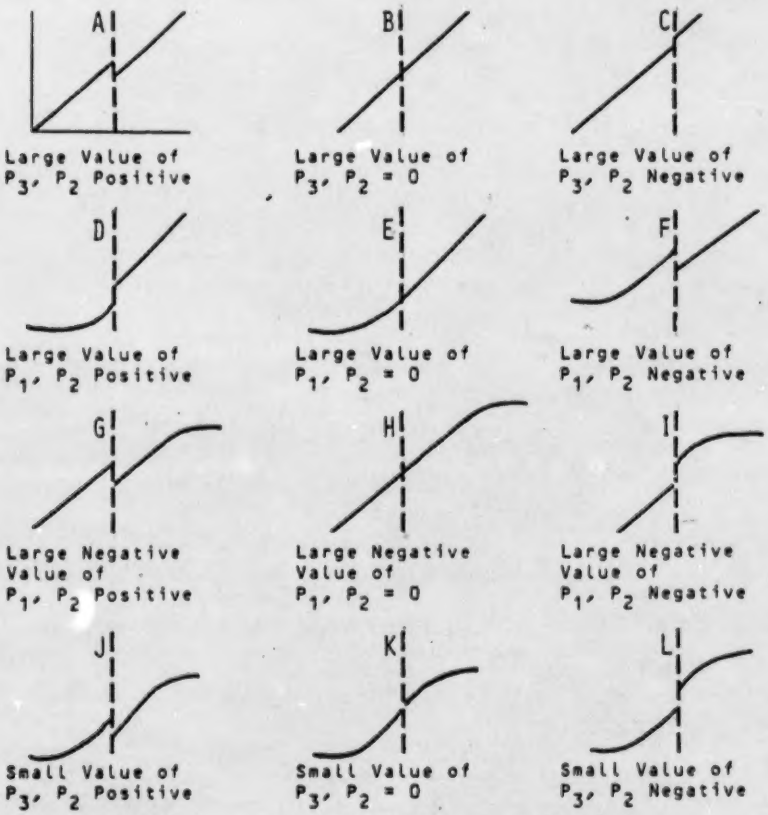
FIGURE 12

FITTING A FOUR-PARAMETER LOGISTIC CURVE PLUS A STEP FUNCTION

THE FUNCTION $\hat{Y} = P_1 + P_2\delta + P_3 \dfrac{e^{P_4(X - P_5)}}{1 + e^{P_4(X - P_5)}}$

WHERE $\delta = 0$ FOR SUBJECTS ABOVE THE CUTOFF, AND
$\delta = 1$ FOR SUBJECTS IN THE TREATMENT GROUP,

GIVES A GENERAL CURVE WITH TREATMENT EFFECT P_2, WHICH CAN ASSUME ANY OF THE FOLLOWING FORMS. THE FUNCTION WAS FIT ITERATIVELY WITH GARY ALLEN'S DERIVATIVE-FREE REGRESSION ROUTINE. STARTING GUESSES WERE $P_1 = P_2 = 0$, $P_3 = 200$, $P_4 = .2$, $P_5 = 50$, $C = 50$.



THUS THIS CURVE FITS FLOOR AND CEILING EFFECTS IF THE POSTTEST DATA DICTATE SUCH CURVILINEARITY, AND GIVES A GOOD ESTIMATE NEAR THE CUT-POINT. LOCAL MINIMA MAY OCCUR, HOWEVER.

Figure 13

Logistic fit: floor effect
Negative treatment effect

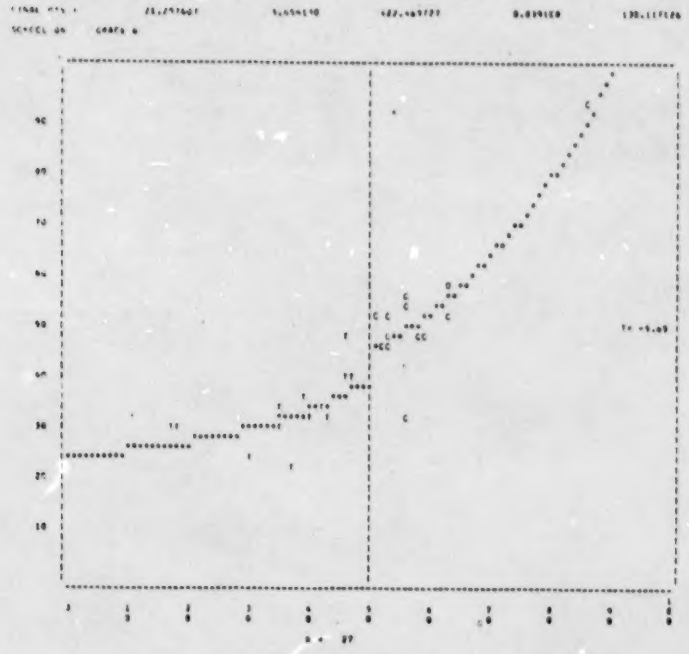


Figure 14

Logistic fit: linear
Negative treatment effect

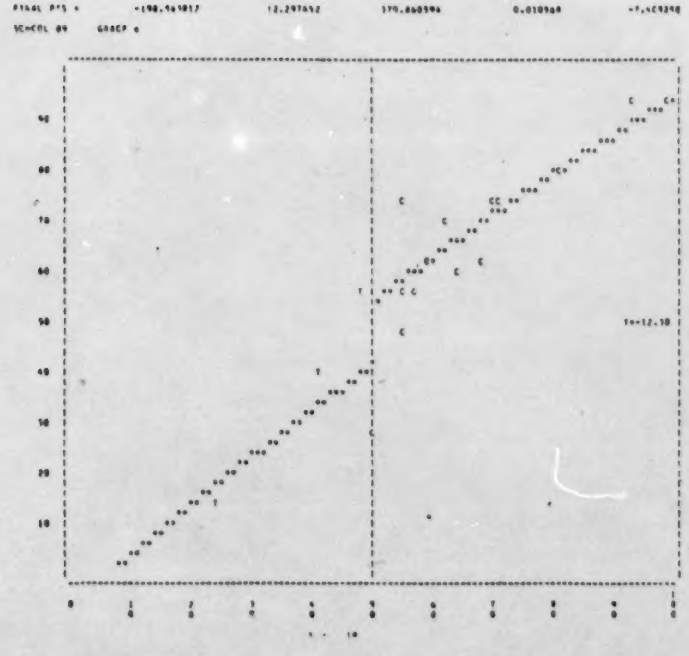
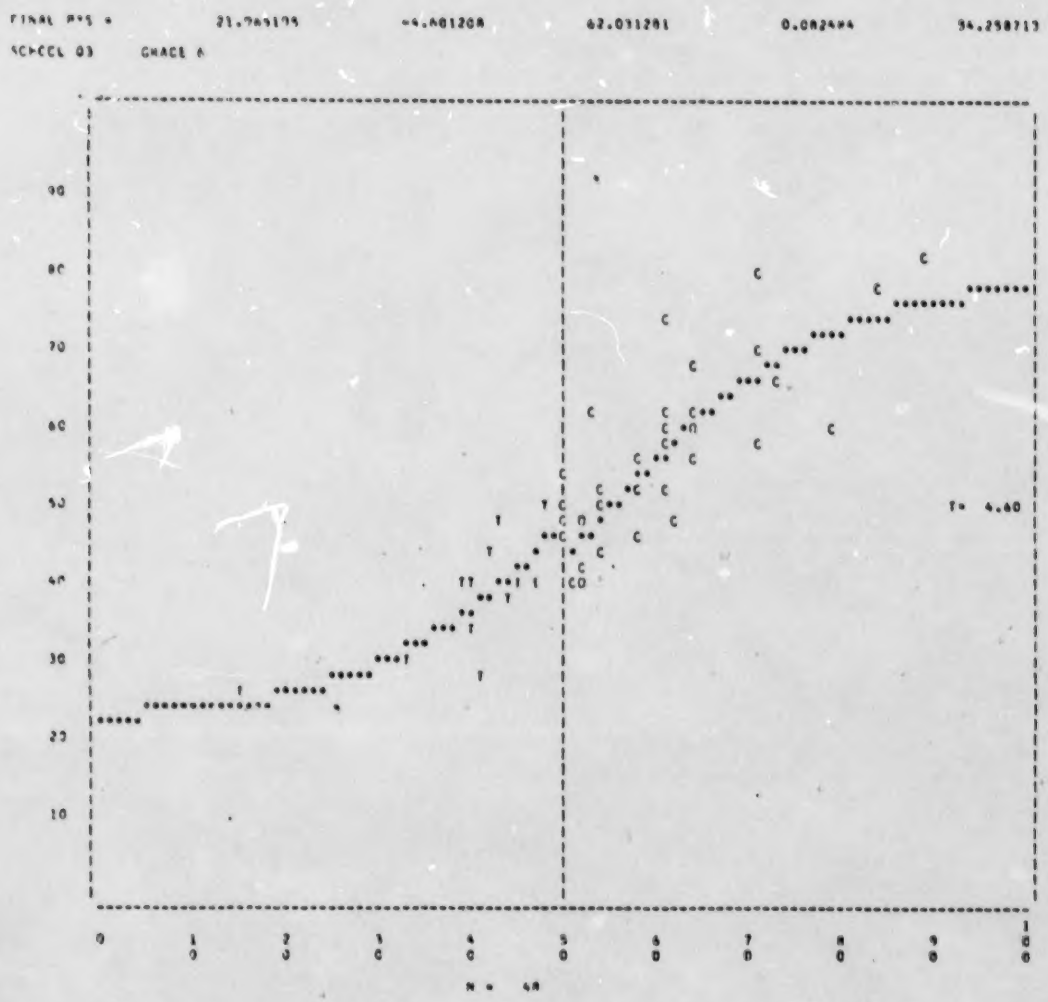


FIGURE 15

Logistic fit: floor and ceiling effects

Positive treatment effect



-2

MODEL C

For comparison purposes, we show the usual model C analysis for each of the three data sets in Figures 16-18. The solid line represents the control group regression function. The dashed line represents the Title I group regression line. Control students are indicated by (*); Title I students are indicated by (+). If the dashed line is generally above the solid line at the lower end of the plot, a positive treatment effect is indicated.

FIGURE 16

Model C Analysis

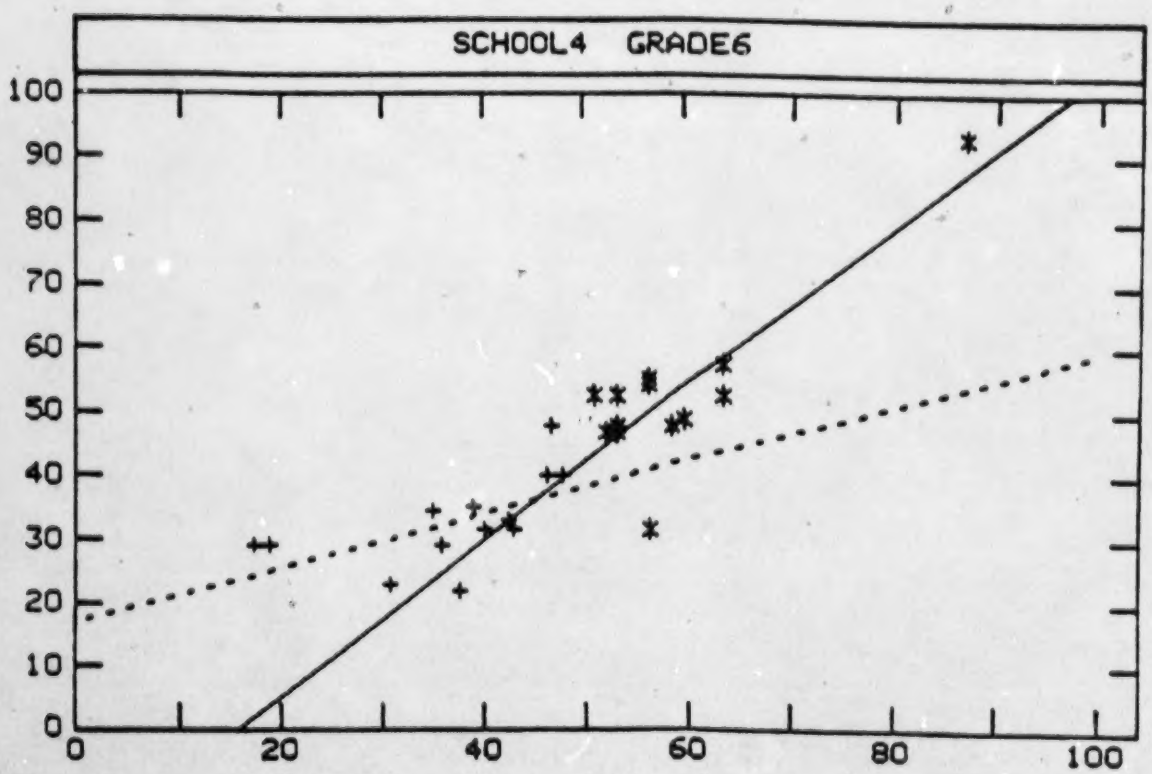


Figure 17

Model C Analysis

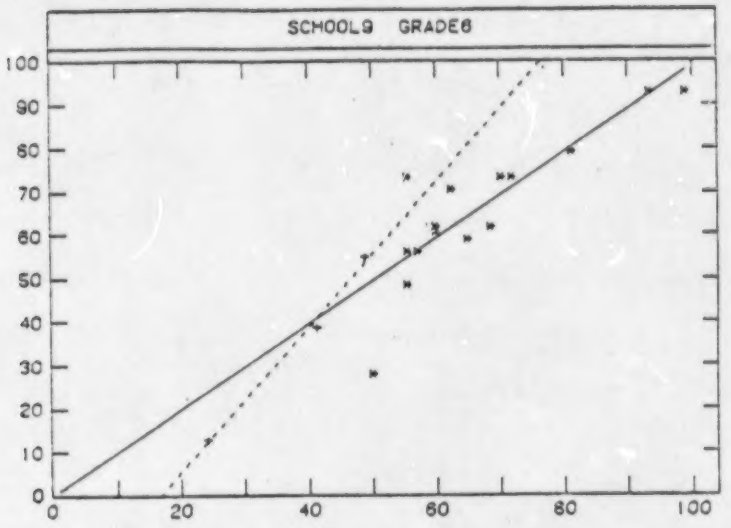
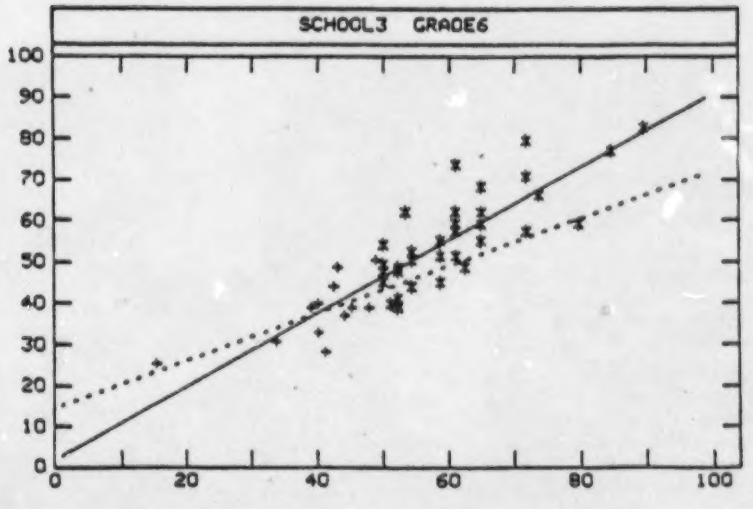


Figure 18

Model C Analysis



Conclusions about model C:

1. Although model C works well when there are no floor or ceiling effects present, nonlinearities may have a large effect on the calculation of impact estimates.

2. In addition to the usual model C analysis, evaluators should apply a parallel-slopes fit and any other fits that seem reasonable, and compare results. If parallel slope and model C procedures give similar estimates, curvilinearity is probably not serious.

3. In any case the interocular impact test (does it hit you between the eyes?) is always advisable. This test requires graphing the scatter plot of pre vs posttest scores and LOOKING.

A suggested reporting format, consisting of treatment and control numbers tested, and of estimated building effects with and without pooling slopes, first using model C, and then using covariance, is given in Figure 19. As was shown in Figure 8, if there is marked curvilinearity these analyses will give different results, alerting the evaluator to look more carefully at the data.

XYZ District Title I Analysis

Grade 3

Building	N T	N C	Extrapolated Building	Extrapolated Pooled	Parallel Lines Building	Parallel Lines Pooled
GRANT	12	20	3.43	3.51	1.02	.51
ADAMS	18	34	-6.21	-.41	1.30	-.4
LEE	8	29	4.53	5.62	3.48	4.21
JEFFERSON	13	42	3.61	4.52	2.91	1.64
Total	51	X	Weighted Average X	X	X	X

REFERENCES

BELSEN, W.A. A technique for studying the effects of a television broadcast, Applied Statistics, Volume 5. 1956. pp. 195-202.

CAMPBELL, Donald T. and Julian C. Stanley. Experimental and quasi-experimental designs. Chicago: Rand-McNally, 1963.

COCHRAN, W.G. The use of covariance in observational studies, Applied Statistics, Volume 18. 1969. pp. 270-275.

DAVIS, F.B. Item Analysis Data: Their compilation, interpretation and use. Harvard Education Papers, No. 2. Harvard Graduate School of Education, Cambridge, Mass., 1946.

MOSTELLER, Frederick and John Tukey. Data Analysis and Regression. Addison Wesley Reading, Mass. 1977.